

Д.О. Половцев

## К ВОПРОСУ О ВАЛИДНОСТИ И НАДЕЖНОСТИ ТЕСТОВ ПРИ ИХ РАЗРАБОТКЕ

**Краткое введение.** Педагогический тест – это система заданий специфической формы, позволяющая качественно оценить структуру и уровень знаний, умений и навыков (уровень обученности) испытуемых. Стандартизованный тест должен иметь спецификацию – паспорт с нормами, условиями тестирования, инструкциями. Если тест должен измерять обученность испытуемого, он должен соответствовать поставленной цели. Такое соответствие называют валидностью, т. е. задания теста должны измерять те навыки и умения, для проверки которых они были составлены. Погрешность тестового измерения определяет его надежность – степень стабильности, устойчивости результатов измерения.

**Целью** данной статьи является рассмотрение валидности и надежности тестов, разрабатываемых для определения уровня сформированности навыков восприятия иноязычной речи на слух выпускников средних общеобразовательных школ.

**Основная часть.** Использование тестовых методик непосредственно связано с исследованием валидности различных приёмов контроля понимания. В зарубежной литературе по вопросам тестирования можно встретить самые разнообразные подходы к определению термина валидности [1]. Валидность теста включает в себя широкий круг понятий, что сказалось, во-первых, в отсутствии единства взглядов среди зарубежных и отечественных тестологов, психологов, методистов на то, что же следует понимать под валидностью, и, во-вторых, в разных вариантах перевода его на русский язык (“адекватность”, “пригодность”, “обоснованность”, “действенность” и др.). Указанное различие во взглядах настолько значительно, что некоторые авторы предпочитают употреблять даже другой термин – “уместность” (relevance), т.е. то, что соответствует данным условиям. Одни педагоги валидность теста определяют по его соответствию изученному материалу (curricular v.), данным математико-статистической обработки его результатов (statistical v.) или логико-психологического анализа (logical and psychological v.).

В самом общем виде валидность может быть определена как важнейшая характеристика теста, которая показывает, что измеряет тест и насколько хорошо он это делает. Иными словами, эта характеристика теста показывает большую или меньшую степень пригодности теста для его использования с определенной целью. Вопросы валидности – это вопросы о том, какие и насколько

правильные и обоснованные выводы могут быть сделаны на основе тестовых оценок, то есть валидность имеет дело с обоснованностью выводов из тестовых оценок или других форм оценивания.

Основным средством валидации (валидация – процесс сбора необходимых данных или оценок для сопоставления их с результатами по тесту, валидность которого хотят определить) педагогических тестов является сопоставление результатов по тестам и нетестовым формам контроля. В этом случае оценки по нетестовым формам контроля выступают в качестве внешнего критерия. Для валидации тестов применяется также и методика рейтинга. Преимущество использования рейтинга для валидации тестов состоит в том, что при рейтинге удаётся получить количественные характеристики, коэффициент корреляции которых с индивидуальными тестовыми результатами может быть легко найден. Этот коэффициент корреляции как раз и является коэффициентом валидности данного теста.

Надёжность, как уже упоминалось выше, является вторым главным показателем эффективности теста. В широком плане под надёжностью понимается «такая сторона предметов, явлений, процессов и систем, которая характеризует степень эффективности и устойчивости становления, функционирования и развития» [2, с. 123]. Что касается психолого-педагогических измерений, то здесь надёжностью мы называем постоянство или устойчивость оценок, получаемых при повторяющихся наблюдениях. Применительно к тестам надёжность определяется постоянством, с которым тест выполняет свою функцию измерительного инструмента. Надёжный тест даёт одни и те же или близкие результаты.

Очевидно, что при разработке теста необходимо учитывать вопросы его валидности и надёжности. Разработка теста требует решения ряда методических задач:

1. Определить цели тестирования (например, текущий контроль, тематический, проверка конечного уровня обученности и др.).
2. Выделить объекты контроля – знания, умения и навыки, которые должны быть проверены. Для этого необходимо провести методический анализ учебных материалов действующих учебников, требований стандарта или программ к обязательным результатам обучения. Перечень и структура объектов контроля зависят от уровня требований, предъявляемых к учащимся: проверяются базовые результаты обучения или умения и навыки более высокого уровня.
3. Разработать структуру теста, то есть расположить выделенные объекты контроля и задания в определённой последовательности – от более простых к более сложным и

определить форму задания. Выбор формы заданий диктуется целью и объектом тестирования. Тестовые задания закрытой формы (множественного выбора, на установление соответствий или последовательности утверждений) могут использоваться для контроля усвоения учащимися языкового материала и для контроля развития рецептивных умений. Продуктивные коммуникативные умения могут проверяться с помощью заданий открытой формы (задания на дополнение или со свободно конструируемым ответом).

3. Отобрать материал для теста, проанализировать типичные ошибки учащихся для подбора дистракторов, если речь идёт о языковых тестах. Особую трудность представляет отбор текстов для проверки коммуникативных умений аудирования. При отборе текстов для аудирования необходимо также учитывать аудитивную трудность текста, включающую в себя паралингвистические и нелингвистические трудности, информационную насыщенность текста.
4. При составлении теста особое внимание необходимо уделить формулированию заданий, а также подбору дистракторов, то есть неправильных ответов в тестах множественного выбора. Дистракторы должны быть правдоподобны, равно привлекательны и по форме, и по содержанию.
5. Провести экспериментальную проверку теста с целью сбора статистического материала для анализа тестовых заданий и ответов к ним. В результате статистического анализа делается вывод о целесообразности их включения в окончательный вариант теста.

**Выводы и рекомендации.** Итак, валидность и надёжность являются одними из главнейших факторов, которые необходимо учитывать разработчикам тестов. Тест может быть надёжным, не будучи валидным, однако он не может быть валидным, не являясь в то же время надёжным. Иными словами, надёжность теста – необходимое условие его валидности.

## **Литература**

1. Рапопорт, И. Тесты в обучении иностранным языкам в средней школе / И. Рапопорт, Г. Сельг, И. Соттер. – Таллин: Валгус, 1987. – 256 с.
2. Пушкин, В. Н. О законе единства надёжности и нарушения / В.Н. Пушкин // Социально-экономические и исторические науки: тезисы докладов XIX Герценовских чтений. – Л., 1966. – С.118–128.